

DATA ANALYSIS IN R

“let the data speak for themselves” -- Tukey

STRENGTHS OF VISUAL DATA ANALYSIS

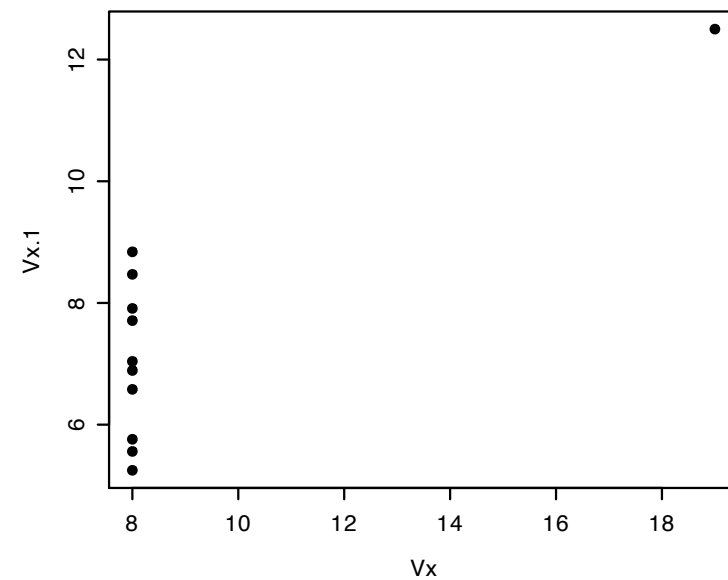
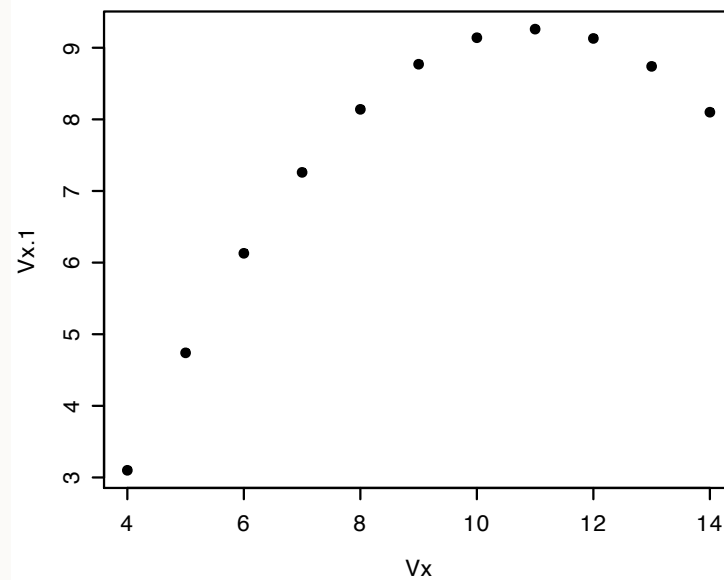
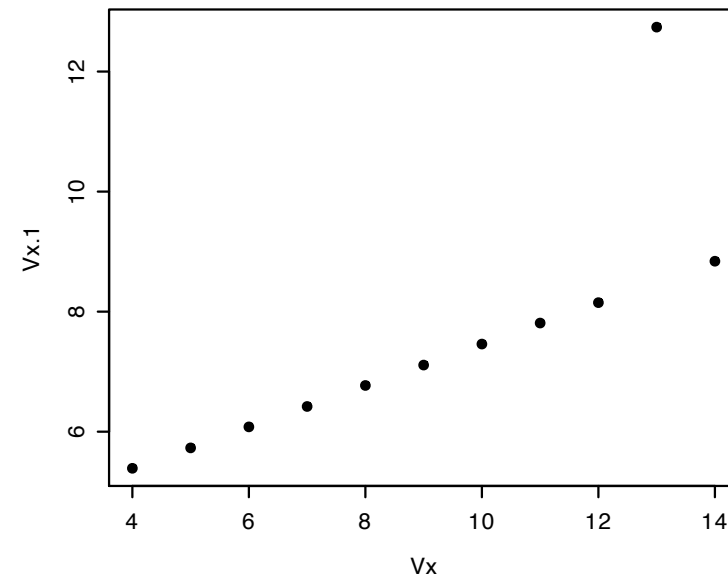
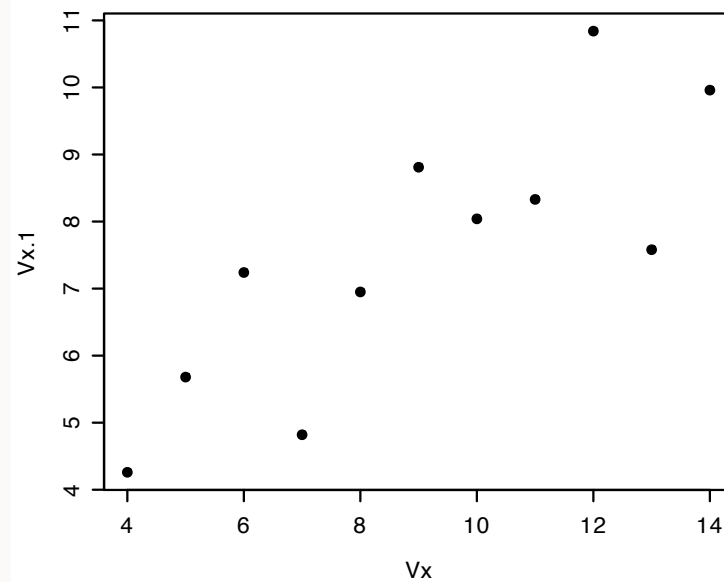
> tab1			> tab2			> tab3			> tab4		
	Vx	Vx.1		Vx	Vx.1		Vx	Vx.1		Vx	Vx.1
1	10	8.04	1	10	9.14	1	10	7.46	1	8	6.58
2	8	6.95	2	8	8.14	2	8	6.77	2	8	5.76
3	13	7.58	3	13	8.74	3	13	12.74	3	8	7.71
4	9	8.81	4	9	8.77	4	9	7.11	4	8	8.84
5	11	8.33	5	11	9.26	5	11	7.81	5	8	8.47
6	14	9.96	6	14	8.10	6	14	8.84	6	8	7.04
7	6	7.24	7	6	6.13	7	6	6.08	7	8	5.25
8	4	4.26	8	4	3.10	8	4	5.39	8	19	12.50
9	12	10.84	9	12	9.13	9	12	8.15	9	8	5.56
10	7	4.82	10	7	7.26	10	7	6.42	10	8	7.91
11	5	5.68	11	5	4.74	11	5	5.73	11	8	6.89

means of x and y same for each data table

$$r=0.82; y=0.5x + 3$$

Anscombe's quartet from Tufte 2001

STRENGTHS OF VISUAL DATA ANALYSIS



REVIEW

- What is R? What is the console? Is R compiled or interpreted?
- What are objects and what *modes* do they fall into?
- What is a data.frame and how can we assign data to it?
- How do we access online help?

OPERATIONS

- $+, -, *, /$
- $\log, \exp, \sqrt{}$
- operations applied to lists / vectors

DESCRIPTIVE STATISTICS

- mean, median, quantile, min, max, var, sd
- summary, fivenum, dnorm
- tapply

PLOTTING

- plot
- hist *or* histogram
- boxplot
- qqplot

EXAMPLE

- subset
- summarize
- plot
- annotate